

# Access to SNOMED Through the National Library of Medicine's Unified Medical Language System

*Kin Wah Fung, MD, MS, MA  
The UMLS Team  
National Library of Medicine  
Bethesda, MD*



# Medical Terminology Standards

- The lack of a common medical terminology standard has been cited as one of the greatest impediments to the widespread development of electronic medical records
- After years of negotiation, an agreement was reached between the National Library of Medicine and the College of American Pathologists to make SNOMED CT available to U.S. users at no cost through the UMLS
- An important step towards establishing a common medical language - a key element in building a unified electronic medical records system



# Challenges of incorporation of SNOMED CT into UMLS

- Complexity and granularity of data structure
- Principle of concept organization
- Size

# SNOMED CT data structure

- Concept-oriented:
  - each concept contains multiple descriptions (concept names)
  - concepts and descriptions have their own unique identifiers and sets of attributes (e.g. ConceptStatus, DescriptionType)
- Relationships:
  - defined at the concept level
  - relationships have their own unique identifiers and set of attributes (e.g. RelationshipGroup)
- Mappings to ICD9CM:
  - in three linked tables (Cross Map Sets, Cross Maps, Cross Map Target)
  - different sets of attributes at each level
- The original release format of the UMLS Metathesaurus cannot represent all the above information

# Original Release Format (ORF) of UMLS Metathesaurus

- Represents information by the **Metathesaurus-Concept-Centric view**
- Main function: as a bridge to bring together various biomedical vocabularies with explicit concept-based connection between terms in one vocabulary and equivalent or related terms in another
- A degree of abstraction and simplification – most of the information is represented at the Concept (CUI) level
- In order to accommodate the level of granularity of SNOMED CT, a new format of the release files is required

# New Rich Release Format (RRF)

- Also supports a new **Source-Centric View** that allows users to retrieve the original information contained in a source vocabulary (source transparency)
- Information is represented at the Atom (AUI) level – an atom is a unit of meaning in a source, usually identifiable by a unique string
- New fields to capture source specific identifiers (e.g.. SCUI, SAUI, SRUI)
- New files to capture mapping information: MRMAP, MRSMAP

# SNOMED CT concept organization

- Concepts organized into 18 top level hierarchies
- Parent-child relationships within hierarchies only (multiple parents allowed) but not across hierarchies
- Other types of relationships between concepts from same or different hierarchies – qualifying or defining relationships
- Use of Description Logic to allow formal representation of the meanings of concepts and their relationships: provides for algorithmic determination of synonymy, redundancy and hierarchical relationships

# Description Logic in SNOMED CT

- Viral pneumonia (disorder) – fully-defined by:
  - Is-a
    - viral respiratory infection (disorder)
    - Infective pneumonia (disorder)
  - Causative agent
    - Virus (organism)
  - Associated morphology
    - Inflammation (morphologic abnormality)
  - Finding site
    - Lung structure (body structure)



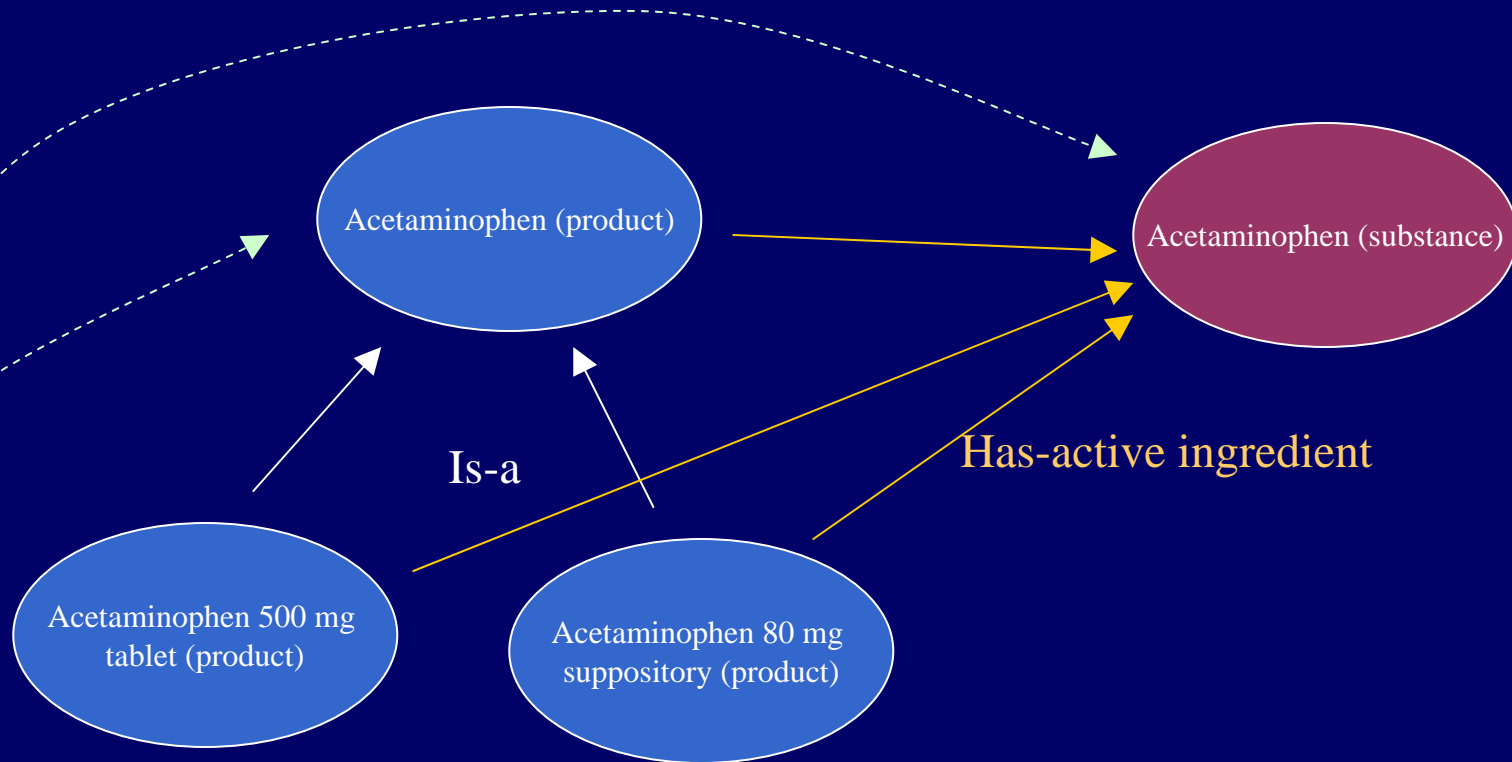
# Different principles of concept organization

- What constitutes a concept?
  - In SNOMED CT:
    - Rigid hierarchies – mutually exclusive
    - The need to define concepts by other concepts based on Description Logic
    - Preference for highest clarity in meaning (fine granularity of concepts)
  - In UMLS:
    - The working principle is ‘useful level of distinction for biomedical professionals (e.g. clinical discourse for clinical terms)’
    - Whether concepts are split depends on common usage
- ⇒ Occasional disagreement in concept organization



**UMLS**

# SNOMED CT Concept View



'This patient is on acetaminophen.'

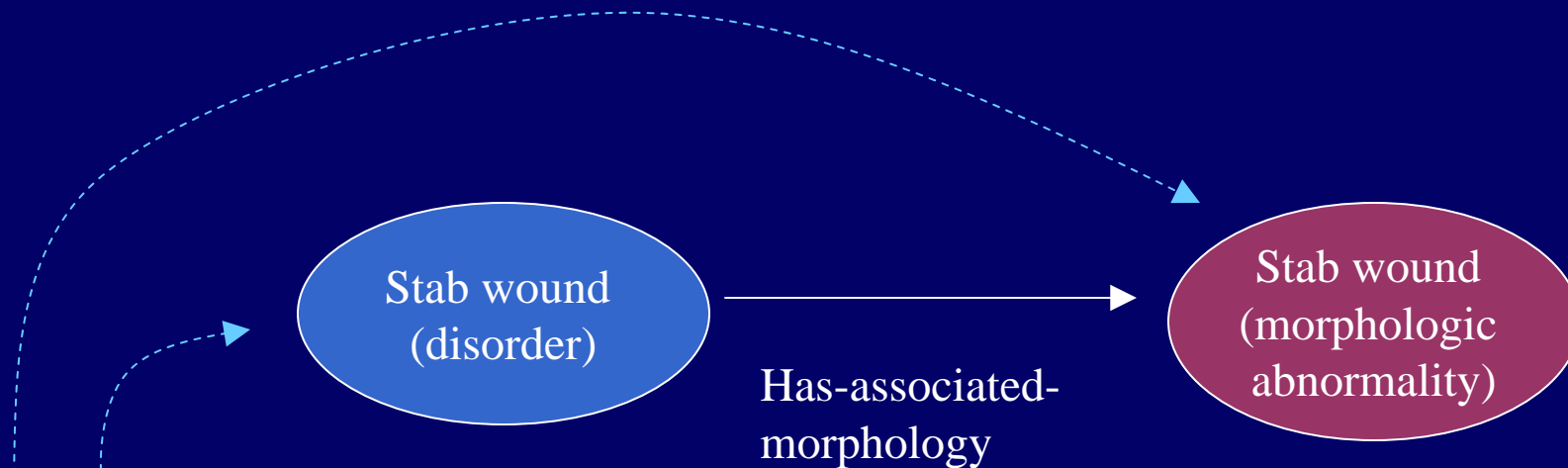
'Tylenol contains acetaminophen.'

Useful clinical distinction?



**UMLS**

# SNOMED CT Concept View



‘The patient was admitted for a stab wound sustained during a fight.’

‘On physical examination, there was a 3 cm stab wound in the right upper quadrant of the abdomen.’

Useful clinical distinction?



**UMLS**

# Other examples of disagreement

- Disorder/finding
  - Change in personality (disorder)
  - Personality change (finding)
- Finding/morphologic abnormality
  - Animal bite (finding)
  - Animal bite (morphologic abnormality)
- Finding/observable entity
  - Antenatal screening finding (finding)
  - Antenatal screening finding (observable entity)

# Source transparency principle

- There is no loss of information in the process of insertion
- Every element of information contained in a source is included in the release files, even though they may be organized differently
- SNOMED CT concepts merged into the same UMLS concept still carry their distinct SNOMED CT ConceptIds and the relationships between them are represented in MRREL
- Proof of source transparency in insertion of SNOMED CT:
  - source-derived files extracted from the original SNOMED CT files
  - UMLS-derived files extracted from the 2004AA release files in the same format as the source-derived files
  - row-by-row comparison proved that they are identical



# Size does matter!

- The insertion of SNOMED CT and the use of the RRF file format will have significant impact on the size of the release files
  - estimated to be about 12 GB compressing to 2.5 GB
- If they continue to be released in CD-ROMs:
  - more and more discs will be needed (already requiring 4 discs before SNOMED CT)
  - problem of big files being split across discs
- Decided to switch to DVD format (4.7 GB/disc)
- Added advantage of installing UMLS directly from the DVD instead of the hard drive, using MetamorphoSys as the installing and customizing program

# SNOMED CT in the 2004AA Release of UMLS

- All current concepts, English descriptions and relationships involving current concepts
  - Concepts: 298,090
  - Descriptions: 736,946
  - Relationships: 1,315,910
- New to the UMLS:
  - Concepts: 37,089
  - Descriptions: 350,464
- Spanish terms, obsolete concepts and descriptions will be available in future releases

# Location of SNOMED CT fields in the Metathesaurus files

- Information contained in every field of SNOMED CT tables is represented in the UMLS Metathesaurus
- For the 3 main SNOMED CT tables: Concepts, Descriptions and Relationships; corresponding fields can be found in 3 RRF files: MRCONSO, MRSAT and MRREL in UMLS (some fields will not be available in ORF files)
- Detailed documentation and sample SQL statements available at:
  - [http://www.nlm.nih.gov/research/umls/Snomed/snomed\\_represented.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_represented.html)



# SNOMED CT Concepts

## SNOMED CT Concepts Table:

ConceptId	ConceptStatus	FullySpecifiedName	CTV3ID	SNOMEDID	IsPrimitive
271737000	0	Anemia (disorder)	XM05A	DC-10009	1

## MRCONSO.RRF:

CUI	...	AUI	...	SCUI	...	SAB	TTY	...	STR	...
C0002871		A3597593		271737000		SNOMEDCT	FN		Anemia (disorder)	
C0002871		A2878480		271737000		SNOMEDCT	PT		Anemia	
C0002871		A2952250		271737000		SNOMEDCT	SY		Absolute anemia	
C0002871		A3095181		271737000		SNOMEDCT	PTGB		Anaemia	
C0002871		A3089808		271737000		SNOMEDCT	SYGB		Absolute anaemia	

## MRSAT.RRF:

CUI	...	SType	Code	...	ATN	SAB	ATV	...
C0002871		SCUI	271737000		CONCEPTSTATUS	SNOMEDCT	0	
C0002871		SCUI	271737000		CTV3ID	SNOMEDCT	XM05A	
C0002871		SCUI	271737000		SNOMEDID	SNOMEDCT	DC-10009	
C0002871		SCUI	271737000		SPRIMITIVE	SNOMEDCT	1	



**UMLS**

# SNOMED CT Descriptions

## SNOMED CT Descriptions Table:

DescriptionId|DescriptionStatus|ConceptId|Term|InitialCapitalStatus|DescriptionType|  
LanguageCode|

406636013|0|271737000|Anemia|0|1|en-US|

## MRCONSO.RRF:

CUI|...|AUI|SAUI|SCUI|...|SAB|TTY|...|STR|...|

C0002871|A2878480|406636013|271737000|SNOMEDCT|PT|Anemia|

## MRSAT.RRF:

CUI|...|METAUI|STYPE|CODE|...|ATN|SAB|ATV|...|

C0002871|A2878480|SAUI|271737000|DESCRIPTIONSTATUS|SNOMEDCT|0|

C0002871|A2878480|SAUI|271737000|INITIALCAPITALSTATUS|SNOMEDCT|0|

C0002871|A2878480|SAUI|271737000|DESCRIPTIONTYPE|SNOMEDCT|1|

C0002871|A2878480|SAUI|271737000|LANGUAGECODE|SNOMEDCT|en-US|



**UMLS**

# Subsetting of the UMLS Metathesaurus

- The Metathesaurus contains concepts and terms from many different source vocabularies designed for different purposes – it must be customized for effective use in most specific applications
- The new RRF format will support easy subsetting of the Metathesaurus
- Content View Flag (CVF)
  - a field in most RRF tables
  - a bit string that defines up to as many content views (subsets) as there are bits in the string
  - a “1” in a particular position indicates that a certain row will be included in that subset
  - expected to be in use from 2004AC release

# The UMLS PHIN Subset

- The CDC is in the process of creating the PHIN Notifiable Condition Mapping Tables to facilitate electronic reporting of notifiable diseases
- The mapping tables include codes from some UMLS source vocabularies:
  - SNOMED CT: 21927003 Bacillus anthracis
  - LOINC: 11469-4 Bacillus anthracis identified : ACnc : Pt : xxx : Ord : Organism specific culture
- A subset of the Metathesaurus can be created using the CVF to identify the codes and terms defined by the PHIN Notifiable Condition Mapping Tables – a handy way for UMLS users to access this list





Center  
Drive

Library  
of  
Medicine

*Thank you!*